INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

How do we perceive the eye intrinsic parameters ?

Thierry Viéville, Jacques Droulez, Chin-Hwee Peh and Alexandre Négri

N° 4030

Octobre 2000

THÈME 3

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--4030--FR+ENG

How do we perceive the eye intrinsic parameters ?

Thierry Viéville*, Jacques Droulez†, Chin-Hwee Peh‡ and Alexandre Négri

Thème 3 — Interaction homme-machine, images, données, connaissances
Projet RobioVis

Rapport de recherche n° 4030 — Octobre 2000 — 39 pages

Abstract: We discuss here how to identify how a visual system deals with the intrinsic calibration parameters of the visual sensor, i.e. the eye, in the case of motion perception.

In computer vision, such an approach is already well developed, from both a geometric and algebraic points of view. It is now well established in this field that calibration mechanisms related to the camera intrinsic parameters allow to adapt such a perceptual system to incoming information and obtain a robust and significant estimation of the structure and motion parameters of the system itself and the scene.

The assumption of this work is that the biological visual perception must also be based on such mechanisms and we have implemented the protocols which allow to verify this hypothesis.

Key-words: Motion perception. Calibration

* INRIA, BP 93, 06902 Sophia, *Thierry.Vieville@inria.fr*
† Collège de France, Paris, *Jacques.Droulez@college-de-france.fr*
‡ National University of Singapore, *elepehch@nus.edu.sg*

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles, B.P. 93, 06902 Sophia Antipolis Cedex (France)
Téléphone : 04 92 38 77 77 - International : +33 4 92 38 77 77 — Fax : 04 92 38 77 65 - International : +33 4 92 38 77 65

Comment percevons nous les paramètres intrinsèques de l'oeil ?

Résumé : On discute ici comment mettre en évidence la façon dont le système visuel gère les paramètres de calibration intrinsèques de l'oeil au niveau de la perception du mouvement.

En vision par ordinateur, cette approche est déjà bien développée, tant du point de vue géométrique qu'algébrique. Il est en effet bien établi que des processus de calibration relatifs aux paramètres intrinsèques de la caméra permettent au système perceptif de s'adapter au mieux au type d'informations qui lui est fourni et d'obtenir à la fois une estimation robuste et significative des paramètres qui caractérisent le mouvement et la structure observée observé.

L'hypothèse de ce travail est que la perpection visuelle biologique doit elle aussi être basée sur de tels mécanismes et son but de mettre en place des protocoles qui puissent le mettre en évidence.

Mots-clés : Perception du mouvement. Calibration

Introduction

Has the brain always a precise *knowledge of the eye intrinsic* (i.e. optical) *parameters*, when performing a perceptual task, such as (see [34, 31] for a review) motion perception ?

This may look like a strange question ! However, if one consider not a biological, but an artificial visual system, this becomes an indeed primary and fundamental question, i.e. do we have to *calibrate* the intrinsic optical parameters of the visual sensors to perform any perceptual task ?

Such *intrinsic* parameters correspond, for instance, to the average focal length of the optical sensor f or the principal point[1] position $\mathbf{m}_0 = (u_0, v_0)$ (i.e. the intersection of the optical axis with the retinal plane, also called "fovea", see Fig. 1). In fact, it has been experimentally demonstrated [13, 38] that, for usual cameras and considering "robotics" visual perception, these three parameters[2] describe properly the projection of a 3D Euclidean point $\mathbf{M} = (X, Y, Z)$ onto a retinal 2D point $\mathbf{m} = (u, v)$ through :

$$\left\{ \begin{array}{rcl} u &=& u_0 + f\,x \\ v &=& v_0 + f\,y \end{array} \right. \text{ with } \begin{array}{rcl} x &=& X/Z \\ y &=& Y/Z \end{array} \qquad (1)$$

considering a frame of reference attached to the retina, which origin is the eye optical centre (if any, hence the image nodal point), the optical axis being aligned with the Z-axis and horizontal and vertical image coordinates being aligned with the X and Y-axes respectively, as illustrated in Fig. 1. The key point here (see [17] for an introduction) is that this knowledge is mandatory to

[1]**Notations:** We write vectors and matrices in bold letters, matrices being written with capital letters. The duals of vectors are represented as the transpose of a vector and scalars in italic, the dot-product being written as $\mathbf{x}^T\,\mathbf{y}$ and the cross-product $\mathbf{x} \times \mathbf{y}$ or $[\mathbf{x}]_\times\,\mathbf{y}$. The identity matrix is written $\mathbf{I}$. We represent the components of a matrix or a vector using subscripts, e.g.: $\mathbf{u} = (u_x, u_y, u_z)^T$. Here $\mathbf{x} \equiv \mathbf{y}$ means $\lambda\,\mathbf{x} = \mathbf{y}$ for some $\lambda \neq 0$.

[2]In computer vision, more "classical" models (e.g. [16]) use two additional "technological" parameters. On one hand (i) the horizontal/vertical scale ratio when pixels are not square but "rectangular", whereas it has been shown [12, 44] that this parameter is always constant in practice, thus can easily be set to one, using a simple coordinate transformation. On the other hand (ii) the skew factor with corresponds [17] to misalignment of the optical axis with respect to the retinal plane, whereas it is, in practice, always negligible for low and high quality cameras [12, 44]. In addition, the 3 parameters model is very efficient with respect to auto-calibration estimation [41, 20].

evaluate Euclidean geometrical quantities such as distances, angles, velocities, etc...

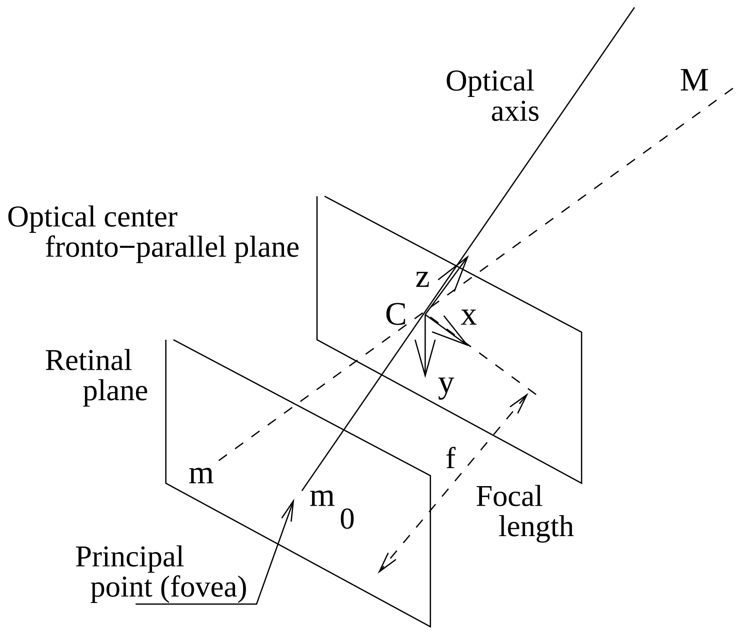

Figure 1: Description of intrinsic parameters of a visual system.

Optical distortions are also to be taken into account [36, 32]. In computer vision, geometric distortions have been modelized, whereas other optical sources of errors (diffraction, achromat, astigmat, etc...) are neglected or taken as a "residual" noise. In fact, such a correction can be easily done using cues such as recti-linearity [6] and treated in a front-end module. They will not be considered here.

Furthermore, taking the eye anatomy into account, it is obvious that such parameters vary from the fovea to the periphery, but also with time and external factors, while their estimation from the direct observation of the environment (very likely its rigid structures) is indeed a solvable problem [28], even for varying and unknown focal length and principal point [20], but numerically ill-conditioned [27], unless active specific displacements are used (mainly rotational displacements [37, 19]).

It is thus a legitimate question to attempt to know whether the brain does not try to "avoid" using such calibration knowledge, either because available visual data may not be sufficient to self-calibrate [18, 27] or just because it is often possible to perform visual or visuo-motor tasks [15, 45, 2] without calibration or with a very approximate knowledge of their values [14]. When not calibrated, a visual system is not observing the Euclidean geometry of the scene, but only its projective or affine geometry, as parameterised in [40, 26] for instance. As a consequence, "calibrating" or not the visual system is equivalent of choosing a given (either Euclidean or projective) geometry to represent spatial and kinematic visual information.

Surprisingly enough, considering for instance perceptual experiments using anamorphozing glasses [11], it has been established that the visual system is able to re-calibrate a Riemannian metric adapted to the glasses deformations, this plastic adaptation allowing a correct perception of the scene Euclidean geometry. The assumption that the brain cognition is more "Euclidean than affine or projective" is reinforced by a recent experiment [29] in which it has been shown that subjects preferentially consider the orthogonality cue (thus an Euclidean cue) rather than parallelism for instance (which is an affine cue) in order to perceive the orientation of a surface drawn using curves.

As a consequence, researches studying either biological motion perception [34, 31] or even the perception of the scene structure from motion cues (e.g. [21, 7]) always assume that the brain uses a calibrated visual system.

Then, the question is not to know if the brain use Euclidean perception, it does, but *how does it manages intrinsic calibration parameters* ? In particular, considering motion perception, we would like to study in this paper, *how does it react to a change of the intrinsic parameters* ?

This is the goal of this paper.

In the next section, we revisit the parameterisation of the retinal projection of the 3D motion of an object in the case of intrinsic parameter variations, in order to be able to formalise our problem.

Then, we discuss how the variations of the intrinsic parameters could be perceived, in the case where either the perceptual system is able to take into account such variables or not.

From this discussion, paradigms are proposed in order to analyse how human cognition is influenced by intrinsic parameters.

Experimental set-up and results are presented and discussed in the last sections.

Parameterising the retinal motion of a rigid object.

Let us revisit the parameterisation of the retinal projection of the 3D motion of an object, as developed in [9, 10] for neuro-physiological models of motion perception, but now considering an un-calibrated visual system as developed for instance in [23].

These developments are based on Taylor expansions of the retinal velocity field, the 3D velocity being assumed to be locally constant and the surface locally planar or quadratic.

We thus parameterise :

- the 3D rigid motion of a point $\mathbf{M} = (X, Y, Z)^T$ through its instantaneous translation vector $\mathbf{T} = (T_x, T_y, T_z)^T$ and rotational velocity $\mathbf{R} = (R_x, R_y, R_z)^T$ so that :
$$\dot{\mathbf{M}} = \mathbf{T} + \mathbf{M} \wedge \mathbf{R}$$
i.e. :
$$\left\{ \begin{array}{l} \dot{X} = T_x + Z\, R_y - Y\, R_z \\ \dot{Y} = T_y + X\, R_z - Z\, R_x \\ \dot{Z} = T_z + Y\, R_x - X\, R_y \end{array} \right. \quad (2)$$

- the local scene structure through the second order Taylor expansion of the proximity, i.e. the inverse of the depth, so that, for instance up to second order, we obtain :
$$P = \frac{1}{Z} = P_z + x\, P_x + y\, P_y + x^2\, P_{xx} + x\, y\, P_{xy} + y^2\, P_{yy} \quad (3)$$
where :
(i) the affine proximity for this retinal point is P_z,
(ii) the 3D Euclidean surface normal is $\mathbf{N} \equiv (P_x, P_y, P_z)^T$,
(iii) while the other quantities (P_{xx}, P_{xy}, P_{yy}) are related to the surface curvature,as schematized in Fig. 2.

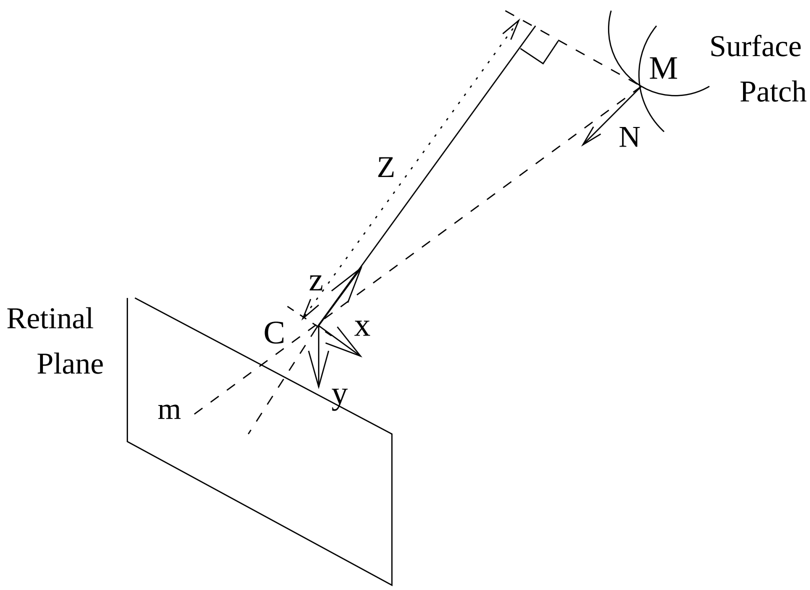

Figure 2: Description of the local scene structure.

From these notations, using the temporal derivation of (1) :

$$\left\{ \begin{array}{lcl} \dot{u} & = & \dot{u}_0 + \dot{f}\,x + f\,\frac{1}{Z}\,(\dot{X} - x\,\dot{Z}) \\ \dot{v} & = & \dot{v}_0 + \dot{f}\,y + f\,\frac{1}{Z}\,(\dot{Y} - y\,\dot{Z}) \end{array} \right. \tag{4}$$

and applying notations introduced in [23] for un-calibrated representation of translation and rotation :

$$\left\{ \begin{array}{lcl} \mathbf{r} & = & \mathbf{A}\,\mathbf{R} \\ \mathbf{t} & = & \mathbf{A}\,\mathbf{T} \end{array} \right. \quad \text{with } \mathbf{A} = \begin{pmatrix} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{pmatrix} \tag{5}$$

using (1), (2), (4) and (5), we can write :

$$\left\{ \begin{array}{lcl} \dot{u} & = & f\left(\dot{\overline{\frac{u_0}{f}}}\right) + \frac{\dot{f}}{f}\,u + \frac{1}{Z}\,(t_x - u\,t_z) + (r_y - v\,r_z) + \frac{u-u_0}{f}\,\epsilon_r \\ \dot{v} & = & f\left(\dot{\overline{\frac{v_0}{f}}}\right) + \frac{\dot{f}}{f}\,v + \frac{1}{Z}\,(t_y - v\,t_z) + (-r_x + u\,r_z) + \frac{v-v_0}{f}\,\epsilon_r \end{array} \right. \tag{6}$$

with : $\epsilon_r = \frac{u-u_0}{f}\,(r_y - v_0\,r_z) - \frac{v-v_0}{f}\,(r_x - u_0\,r_z)$.

Considering also the local structure of the scene, as parameterised in (3) the expression of the second order expansion of the retinal velocity field may be written :

$$\left\{ \begin{array}{lllllllllllll} \dot{u} & \simeq & \bar{u}_0 & + & \bar{u}_1 \, u & + & \bar{u}_2 \, v & + & \bar{u}_{11} \, u^2 & + & \bar{u}_{12} \, u \, v & + & \bar{u}_{22} \, v^2 \\ \dot{v} & \simeq & \bar{v}_0 & + & \bar{v}_1 \, u & + & \bar{v}_2 \, v & + & \bar{v}_{11} \, u^2 & + & \bar{v}_{12} \, u \, v & + & \bar{v}_{22} \, v^2 \end{array} \right. \tag{7}$$

while the proximity, from (3) and (1), is given in a retinal frame of reference, i.e. :

$$P = \frac{1}{Z} = p_z + u \, p_u + v \, p_v + u^2 \, p_{uu} + u \, v \, p_{uv} + v^2 \, p_{vv} \tag{8}$$

with :

$$\left\{ \begin{array}{llllllll} p_{uu} & = & P_{xx}/f^2 & , & p_u & = & P_x/f & - (2 \, u_0 \, p_{uu} + v_0 \, p_{uv}) \\ p_{uv} & = & P_{xy}/f^2 & , & p_v & = & P_y/f & - (u_0 \, p_{uv} + 2 \, v_0 \, p_{vv}) \\ p_{vv} & = & P_{yy}/f^2 & , & p_z & = & P_z & - (u_0 \, p_u + v_0 \, p_v + u_0^2 \, p_{uu} + u_0 \, v_0 \, p_{uv} + v_0^2 \, p_{vv}) \end{array} \right.$$

while, after some algebra and using (1), (2), (8), (4) and (5), we obtain : +

$$\left\{ \begin{array}{llllllll} \bar{u}_0 & = & f \, \dot{\left(\frac{u_0}{f} \right)} & + & p_z \, t_x + r_y & & & - \;\; u_0 \, (v_0 \, r_x - u_0 \, r_y)/f^2 \\ \bar{u}_1 & = & \frac{\dot{f}}{f} & - & p_z \, t_z + p_u \, t_x + & & & + \;\; (v_0 \, r_x - 2 \, u_0 \, r_y + u_0 \, v_0 \, r_z)/f^2 \\ \bar{u}_2 & = & & & p_v \, t_x - r_z & & & + \;\; u_0 \, (r_x - u_0 \, r_z)/f^2 \\ \bar{u}_{11} & = & & & -p_u \, t_z & + & p_{uu} \, t_x & + \;\; (r_y - v_0 \, r_z)/f^2 \\ \bar{u}_{12} & = & & & -p_v \, t_z & + & p_{uv} \, t_x & - \;\; (r_x - u_0 \, r_z)/f^2 \\ \bar{u}_{22} & = & & & & + & p_{vv} \, t_x & \end{array} \right. \tag{9}$$

and :

$$\left\{ \begin{array}{llllllll} \bar{v}_0 & = & f \, \dot{\left(\frac{v_0}{f} \right)} & + & p_z \, t_y - r_x & & & - \;\; v_0 \, (v_0 \, r_x - u_0 \, r_y)/f^2 \\ \bar{v}_1 & = & & & p_u \, t_y + r_z & & & - \;\; v_0 \, (r_y - v_0 \, r_z)/f^2 \\ \bar{v}_2 & = & \frac{\dot{f}}{f} & - & p_z \, t_z + p_v \, t_y & & & + \;\; (2 \, v_0 \, r_x - u_0 \, r_y - u_0 \, v_0 \, r_z)/f^2 \\ \bar{v}_{11} & = & & & & + & p_{uu} \, t_y & \\ \bar{v}_{12} & = & & & -p_u \, t_z & + & p_{uv} \, t_y & + \;\; (r_y - v_0 \, r_z)/f^2 \\ \bar{v}_{22} & = & & & -p_v \, t_z & + & p_{vv} \, t_y & - \;\; (r_x - u_0 \, r_z)/f^2 \end{array} \right. \tag{10}$$

These equations describe the retinal displacement, i.e. what can be locally perceived (up to the second order) on the retina when observing a rigid object in motion.

Note that in the "calibrated case" the related equations correspond to particular case of those with $u_0 = v_0 = 0$, $f = 1$, so that the retinal coordinates (u, v) correspond to (x, y), with $\mathbf{r} = \mathbf{R}$ and $\mathbf{t} = \mathbf{T}$.

Let us analyse these equations.

What can be perceived in the un-calibrated case ?

In the general case, from (9) and (10), we have 12 equations, but to be related to $(u_0, v_0, f, \dot{u}_0, \dot{v}_0, \dot{f})$, $\mathbf{r}$, $\mathbf{t}$ and $(p_z, p_u, p_v, p_{uu}, p_{uv}, p_{vv})$ i.e. 18 unknowns. We thus are in an algebraic situation with "too many unknowns" and may :

- use higher-order equations [1], when considering second-order equations (with velocities and accelerations, i.e. using at least three views) the motion parameters become "observable";

  we will experiment if the motion perception could be based on such higher-order constraints, but usual models in biological vision are not based on such cues because they are not clearly observed in neurophysiology [34, 31];

- directly use the "observable" i.e. the motion field parameters and thus, as introduced by [30] and [15], only consider the affine or projective geometrical properties of the scene , this is called "weak-calibration",

- introduce additional constraints (constant calibration parameters, knowledge of the "horizon" i.e. the plane at infinity, particular displacements, etc...) in order to be able to easily analyse these parameters, that is consider specific situations (as synthesised in [41]), this is called "self-calibration",

as developed now.

Foveal motion perception

From equation (6), as discussed in [23], we may first consider motion perception close to the principal point, i.e. the foveal perception.

By "foveal" perception, we mean that the term with ϵ_r is negligible since it is related to second order terms in $((u - u_0)/f, (v - v_0)/f)$.

The "focus of expansion" and the "retinal rotation" If we consider the *projections* of the translation $\mathbf{t}$ (also known[3] as the "focus of expansion") and of the rotation $\mathbf{r}$, these equations, for fixed intrinsic parameters, have exactly the same form as in the calibrated case. This is illustrated in Fig. 3. This is not true, at the "second order", i.e. when the term ϵ_r is no more negligible, as visible in (9) and (10).

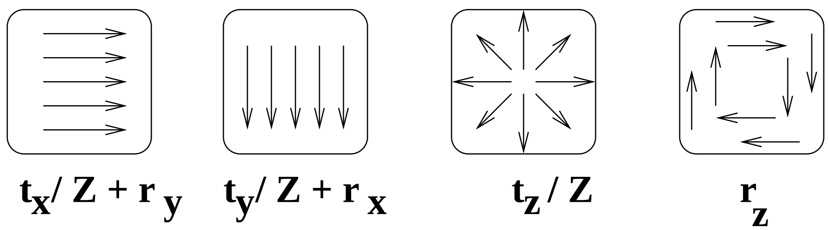

Figure 3: Aspects of the retinal field close to the fovea.

In particular it is obvious that we easily distinguish t_z and r_z but neither t_x from r_y nor t_y from r_x.

In terms of cognition, Droulez [9], suggests a perceptual model for the brain to get rid of this ambiguity. Computer vision algorithms use second order terms to disambiguate (e.g. [43]).

The "motion equation" If we eliminate $\frac{1}{Z}$ from (6) we obtain a constraint on the calibration and motion parameters with may be written :

$$\dot{\underline{\mathbf{m}}}^T \, [\mathbf{t}]_\times \, \underline{\mathbf{m}} + \underline{\mathbf{m}}^T \, \mathbf{S} \, \underline{\mathbf{m}} = 0 \tag{11}$$

with $\underline{\mathbf{m}} = (u, v, 1)^T$, $[\mathbf{t}]_\times \, \mathbf{x} = \mathbf{t} \times \mathbf{x}$, while : $\mathbf{t}^T \, \mathbf{S} \, \mathbf{t} = 0$.

This corresponds to continuous form [39] of the "fundamental motion equation" of Longuet-Higgins [24] and Maybank-Faugeras [28].

[3] The retinal point corresponding to the projection of the translation (i.e. the epipole [17]) corresponds to the retinal location of the vanishing point for the direction of the translation.

In the foveal case (see [39] for a general expression) we obtain :

$$\mathbf{S} = \begin{pmatrix} -2\,a & 0 & b \\ 0 & -2\,a & c \\ b & c & -2\,d \end{pmatrix} \text{ with } \left\{ \begin{array}{lcl} a & = & r_z\,t_z \\ b & = & (r_x - f\,\dot{\left(\frac{v_0}{f}\right)})\,t_z + r_z\,t_x - \frac{\dot{f}}{f}\,t_y \\ c & = & (r_y + f\,\dot{\left(\frac{u_0}{f}\right)})\,t_z + r_z\,t_y + \frac{\dot{f}}{f}\,t_x \\ \\ d & = & (t_x\,b + t_y\,c - (t_x^2 + t_y^2)/t_z\,a)/t_z \end{array} \right. \tag{12}$$

so that the matrix $\mathbf{S}$ is a function of $\mathbf{t}$ (with two degrees of freedom since $\mathbf{t}$ is only defined up to a scale factor) and of three other independent quantities (a, b, c), thus function of 5 independent parameters.

As a consequence, for foveal motion perception we only need 5 generic points as for the calibrated case (e.g. [17]) whereas 7 generic points are required for larger fields of views (see [39] for a development in the continuous case).

From this foveal motion equation, we easily compute :
(i) the focus of expansion $\mathbf{t}$,
(ii) if the intrinsic parameters are constant and $t_z \neq 0$ the projection of the rotation $\mathbf{r}$ (while if $t_z = 0$ we only recover r_z).

However, we :
(iii) never can compute the intrinsic parameters (they only appear in extra-foveal motion terms, which may explain the numerical difficulty to estimate them in practice (e.g. [27])) and we also
(iv) never can compute their variations (i.e. track them from one image to another) since these three variations $(f\,\dot{\left(\frac{u_0}{f}\right)}, f\,\dot{\left(\frac{v_0}{f}\right)}, \frac{\dot{f}}{f})$ are only constrained by two equations.

Recovery of structure and motion Since it is not possible to effectively solve the motion equations in the general case by eliminating $\frac{1}{Z}$, we may use the parameterisation of the scene structure and solve (9) and (10).

In this case, as visible in (9) and (10) and made explicit in appendix A, we can eliminate :
(1) the structural parameters p_{uu}, p_{uv}, p_{vv}, p_u, p_v and $(p_z\,t_z - \frac{\dot{f}}{f})$ and,

(2) the rotational parameters $(r_x - \dot{\left(\frac{v_0}{f}\right)})$, $(r_y + \dot{\left(\frac{u_0}{f}\right)})$ and r_z, mixed with intrinsic parameters variation, since they operate linearly in the equations, with coefficients only function of the translation $\mathbf{t} = (t_x, t_y, t_z)$.

As a consequence, we also do not recover the intrinsic parameters in this case, whereas *their variation is perceived as rotations or translation in depth* as discussed in the sequel.

These equations are singular only if $t_x = t_y = 0$ but in such a case, we do not recover p_{uu}, p_{uv}, p_{vv} but still obtain p_u, p_v, $(p_z\, t_z - \frac{\dot{f}}{f})$ and $(r_x - \dot{\left(\frac{v_0}{f}\right)})$, $(r_y + \dot{\left(\frac{u_0}{f}\right)})$, r_z, whereas t_z is no more evaluated in this case.

The parameters of translation may be obtained from the motion equation (11) as discussed previously, but we show in appendix A that they may be also computed directly from the same equations since two linear constraints exists between their components defining the translation up to a magnitude factor.

Relation with the case of a pure translation or a zoom In the case where we have a (i) pure translation or a (ii) camera focal-length variation, i.e. a zoom, (that is a variation of the intrinsic parameters, plus a translation of the optical centre) we also have $\epsilon_r = 0$. As a consequence, these *are particular cases of foveal perception.*

In the motion equations $\mathbf{S} = 0$ for a pure translation, while since $a = 0$ in (12), $\mathbf{S}$ has a simpler form for a zoom.

Relation with the case of a "retinal invariant rotation" In the case where $R_x = R_y = 0$ i.e. the rotation has been cancelled except in the plane of the retina, we still are in the same situation as a foveal perception, but with $r_x = u_0\, R_z$ and $r_y = v_0\, R_z$. This means that if $R_z \neq 0$ we obtain particular equations to estimate u_0 and v_0 if they are constant.

As a consequence, this is also *a particular cases of foveal perception.*

This had been already made explicit in a more specific situation by [23].

Considering the affine component of the retinal field Considering "foveal" perception means using a rather small visual field. In such a case, the numerical stability of the estimation of second order terms in (9) and (10) (i.e. $\bar{u}_{11}$, $\bar{u}_{12}$, $\bar{u}_{22}$, $\bar{v}_{11}$, $\bar{v}_{12}$, $\bar{v}_{22}$) may not be sufficient.

As a consequence, we must analyse what could be perceived if we analyse motion perception using only the 6 coefficients (i.e. $\bar{u}_0$, $\bar{u}_1$, $\bar{u}_2$, $\bar{v}_0$, $\bar{v}_1$, $\bar{v}_2$) of the affine part of the retinal field.

In other words small visual fields are to be modelized by affine retinal motion fields [5, 23]

In this case, we do not recover p_{uu}, p_{uv}, p_{vv} but still obtain p_u, p_v, $(p_z\, t_z - \frac{\dot{f}}{f})$ and $(r_x - \dot{\left(\frac{v_0}{f}\right)})$, $(r_y + \dot{\left(\frac{u_0}{f}\right)})$, r_z, since, in the equations made explicit in appendix A, we only use affine coefficients of the retinal field to compute these quantities.

Furthermore, if it is not possible to estimate the translation from these equations, it can still be estimated, using the motion equation.

Therefore, using small visual fields does not restrain what could be recovered from (9) and (10).

However, if $t_x = t_y = 0$, p_u and p_v are no more observable. This is thus a difference with motion perception using second-order terms. In particular if one generates a stimulus of a slanted plane undergoing a translation along the z-axis only, according to this model, the plane orientation is observable if and only if the visual does not only use affine components of the retinal field. This may complete existing comparisons for structure and motion perception between large and small visual field, already experimented in biological vision for instance by [8], while a recent work [9, 10] already demonstrates that human motion perception is mainly related to this affine components.

Self-calibration of the intrinsic parameters

Calibration from the motion constraints Intrinsic parameters estimation from the motion constraints (i.e. eliminating depth in the equations) is indeed possible [28] even for varying and unknown focal length and principal point [20] (although in that case at least 4 images, or equivalently third order temporal derivatives, are needed), but numerically ill-conditioned [27] so that, often, higher-order constraints [3] are required.

In our context, it is clear that such self-calibration mechanisms may be not be considered as plausible model to explain how the brain may estimate eye intrinsic parameters : extra-foveal motion perception is required but in such a case optical properties are far from being linear, higher-order parameterisations of the motion field are mandatory but it seems that first order parameterisations of the motion field are mainly taken into account as cognitive cues [9, 10], as now done in computer vision [4].

Calibration with pure rotations However, when using rotational displacements, the calibration of the intrinsic parameters becomes a much more simple task. When performing a pure rotation [19] or even more generally a fixed axis rotation [37], self-calibration becomes a tractable task. More specifically using our 3 parameters calibration models and pure rotation explicit linear equations can be made explicit [41].

In any case, it must be known from an external source of information that this is a pure rotation, since a pure rotation can not be distinguished from the observation of a single planar structure (see [42] for a discussion).

In the present contribution we have verified that considering the present retinal motion parameterisation, as made explicit in appendix A :

- if the intrinsic parameters are constant during the rotation we obtain a direct estimation of the intrinsic parameters from a set of redundant linear equations,
- if the intrinsic parameters vary we obtain a direct estimation of the principal point location and focal length variation, whereas the principal point variation and the focal length value can not be estimated independently,

while in both cases we obtain a direct estimation of the rotation, knowing the intrinsic parameters.

In practice, this means that we can directly estimate the intrinsic parameters during a rotation if the principal point location is fixed while the focal length may vary. We could use the following algorithm :

$$\left\{ \begin{array}{rcl} \alpha &=& (\bar{u}_{12} + \bar{v}_{22})/2 \\ \beta &=& (\bar{u}_{11} + \bar{v}_{21})/2 \\ u_0 &=& -(\alpha\,(\bar{u}_2 + \bar{v}_1) - \beta\,(\bar{v}_2 - \bar{u}_1))/(\alpha^2 + \beta^2) \\ v_0 &=& -(\beta\,(\bar{u}_2 + \bar{v}_1) + \alpha\,(\bar{v}_2 - \bar{u}_1))/(\alpha^2 + \beta^2) \\ \delta &=& 3\,((\bar{v}_2 + \bar{u}_1) + (\beta\, u_0 + \alpha\, v_0)/2) \\ \epsilon_u &=& \bar{u}_0 - (\bar{v}_2 - 2\,\bar{u}_1 + \delta)/3\, u_0 + \bar{u}_2\, v_0 \\ \epsilon_v &=& \bar{v}_0 + (2\,\bar{v}_2 - \bar{u}_1 - \delta)/3\, v_0 + \bar{v}_1\, u_0 \\ f &=& (\alpha\,\epsilon_u + \beta\,\epsilon_v)/(\alpha^2 + \beta^2) \\ \epsilon &=& [\epsilon_u - \beta\, f^2, \epsilon_v - \alpha\, f^2] \end{array} \right. \tag{13}$$

to compute u_0, v_0 and f. Here ϵ is an indicator of the principal point variation (although not a real estimation since it is expected fixed), in fact of the estimation error. This algorithm is unbiased with respect to focal length variations (estimated as δ), contrary to what is proposed in [19] or [41]. In any case, this algorithm is well defined if and only if $R_x \neq 0$ or $R_y \neq 0$.

If we only consider the 6 equations of the affine part of the retinal field, we still can easily eliminate the 3 components of the rotation and obtain by linear combinations of the equations 3 equations between the intrinsic calibration parameters and their variations, as also made explicit in appendix A. These three algebraic homogeneous equations of degree two have two solutions with $f = 0$ and two solutions which may correspond to the physical intrinsic parameters.

When calibration is not needed ... the planar case example. Finally, several aspect of motion perception do not require calibration [15, 45, 2].

One typical example is the "planar case", i.e. the motion perception of a single planar surface. A planar surface, even with constant intrinsic parameters, yields 12 unknowns but only 8 of the previous equations are independent [41]. The Euclidean location of the plane and its motion is thus not entirely recoverable.

More precisely, let us consider a plane P parameterised by its distance to the origin $d(O, P) = 1/\sqrt{P_x^2 + P_y^2 + P_z^2}$ and its normal $\mathbf{N} = d(O, P)\,(P_x, P_y, P_z)$

with $||\mathbf{N}|| = 1$ so that, from (3) and (8) :

$$M \in P \Leftrightarrow \mathbf{N}^T M = d(O,P) \Leftrightarrow -1 + P_x X + P_y Y + P_z Z = 0 \Leftrightarrow \frac{1}{Z} = p_u u + p_v v + p_z$$

The related retinal displacement, say $(\dot{u}_P, \dot{v}_P)$, has the form given in (7) but with $\bar{u}_{22} = 0$, $\bar{v}_{11} = 0$, $\bar{u}_{12} = \bar{v}_{22}$ and $\bar{u}_{11} = \bar{v}_{12}$. In appendix A, these equations have been made explicit. We show that they are not solvable in general case. However, as initiated in [23], considering "retinal invariant rotation" (i.e. $R_x = R_y = 0$) they are solvable in (p_u, p_v, t_x, t_y) up to a scale factor if and only if either :

1. we have $R_z = 0$, that is if we totally cancelled the rotation;

   in that case we obtain an estimation of the time-to-collision (see (14)) $\frac{\dot{f}}{f} - \frac{1}{Z} T_z = 0$, using the affine part of the retinal field. From its second order part we estimate T_z, unless we observe a fronto-parallel plane, and then obtain equations for the principal point variations;

   or
2. we have $\frac{\dot{f}}{f} - \frac{1}{Z} T_z = 0$ that is if we have a pure "retinal displacement" i.e. the rigid displacement left the retinal plane invariant or is compensated by a focal length variation;

   in that case we cannot estimate p_z and eliminating its value, find a unique equation with respect to the principal point location (unless $R_z = 0$) and variations.

This generalises what had been discussed in [23].

Here, there is a symmetry in the equations so that if (p_u, p_v, t_x, t_y) is a valid solution, (t_x, t_y, p_u, p_v) is also a valid solution of the same equations. With respect to this "orientation-translation" duality, it has been shown [9, 10] that the human visual system always choose a solution corresponding to the plane with tends to be the more fronto-parallel.

Not trying to recover these Euclidean quantities, it is however very easy (e.g. [33]) to recover directly from the observable, the position of a point M with respect to the observed plane P.

A point is in front of a plane (say $M > P$) if and only if $\frac{\mathbf{N}}{||\mathbf{N}||}^T M > d(O, P)$ and we easily see that :

$$M > P \Leftrightarrow \delta = (p_z + p_u\, u + p_v\, v) - \frac{1}{Z} = \frac{d(M, P)}{d(O, P)\, (\mathbf{z}^T M)} > 0$$

as derived in [42], while from (4) we can write :

$$\left\{ \begin{array}{rcl} \dot{u} & = & \dot{u}_P + \delta\, (t_x - x\, t_z) \\ \dot{v} & = & \dot{v}_P + \delta\, (t_y - y\, t_z) \end{array} \right.$$

We thus very easily compute, unless $\mathbf{t} = 0$, the relative weighted position δ of a point with respect to a plane.

Experimenting with "un-calibrated" motion perception

Introducing intrinsic parameters variations.

At this stage, we would like to experiment how cognition deals with the calibration parameters. Our "computational" assumption is that the brain should rather use the previous equations that equations which omit to consider the intrinsic calibration parameters.

How to make the difference ? The general idea is to propose stimuli in which in make the intrinsic parameters vary, with the idea that they will *perturb* the perception of a system which ignore such parameters.

This general idea is corroborated by the fact that the previous equations mainly shows that the introduction of constant calibration parameters mainly induces a simple "change of coordinates" but qualitatively do not change the equations. This is in particular the case for foveal motion perception.

A step further, intrinsic parameters variation is twofold and always perceived in interaction with other parameters as follows :

Focal length variation versus translation in depth A variation of the focal length f (i.e. a zoom) is in direct interaction with the translation along the optical axis t_z since they act in the equation through a common term (as visible in (6)) :

$$\frac{1}{\tau_c} = \frac{\dot{f}}{f} - \frac{1}{Z} t_z \tag{14}$$

known as time-to-collision (i.e. the time, when no zoom, for a constant velocity fronto-parallel planar target, to cross the optical centre fronto-parallel plane $(C, \mathbf{X}, \mathbf{Y})$, defined in Fig. 1).

In other words, *a zoom is equivalent of the depth translation of a fronto-parallel plane.*

Considering our model, $\tau_c^{-1} = \frac{\dot{f}}{f} - p_z\, t_z$ and algebraic developments reported in appendix A demonstrate that:
- in the precise case of a pure translation with $t_x \neq 0$ or $t_y \neq 0$ i.e. (i) no rotation ($R_x = R_y = R_z = 0$) and (ii) no variation of the principal point ($\dot{u}_0 = \dot{v}_0 = 0$) we obtain a direct estimation of $p_z = \frac{\bar{u}_0\, t_x + \bar{v}_0\, t_y}{t_x^2 + t_y^2}$,
- this is still the case if $R_x = R_y = \dot{u}_0 = 0$ and $t_x \neq 0$ (here $p_z = \bar{u}_0 / t_x$) or if $R_x = R_y = \dot{v}_0 = 0$ and $t_y \neq 0$ (here $p_z = \bar{v}_0 / t_y$),
- whereas in any other generic cases p_z can not be estimated independently from τ_c^{-1} so that its estimation is biased if there is a focal length variation, more precisely an estimation $\tilde{p}_z$ of p_z in which $\frac{\dot{f}}{f}$ is neglected is given by :

$$\tilde{p}_z = p_z \left[1 - \frac{\dot{f}}{f} \frac{1}{t_z} \right] \tag{15}$$

Principal point variation versus rotation A variation of the principal point induces a retinal translation through a term :

$$\mathbf{t}_\rho = \begin{pmatrix} f \dot{\left(\frac{u_0}{f}\right)} & + & r_y & + & p_z\, t_x \\ f \dot{\left(\frac{v_0}{f}\right)} & - & r_x & + & p_z\, t_y \end{pmatrix} \tag{16}$$

which combines translations and rotations.

In particular, it has been shown here that *for foveal perception principal point variations are perceived as rotations.* In other paradigms, (e.g. pure rotations) principal point variations is also not computable.

As a consequence, the main parameter to consider is **focal length variation**, i.e. zoom. In terms of perception, a focal length variation corresponds to a *magnification* or *enlargement* of the perceived object, as using a telescope. Here we have to experiment whether this will be discriminated from an *approach* of the object. As not being a "natural" (the eye has no zoom !) subjects should be trained before the experiment with another visual indication (e.g. a background) about the fact they perceived a *magnification* of the scene or an *approach* of the observed object.

If the brain does *not* model intrinsic parameters it will not, otherwise it will.

Using planar slant estimation to test what is perceived.

Following [], we consider a slanted plane (either fronto-parallel or slanted around an horizontal axis) in translation and propose to look at the estimation of the slant angle, as schematized in Fig. 4.

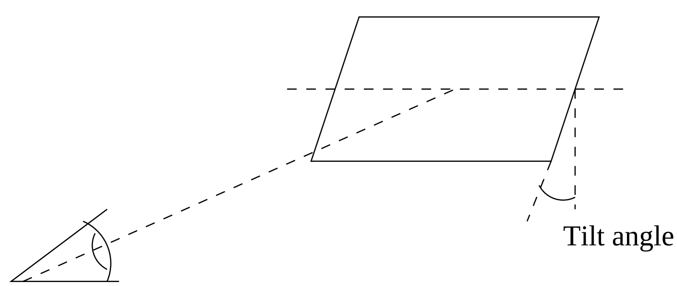

Figure 4: Schematized representation of the chosen experimental stimulus.

From our model, the plane slant angle θ may be estimated from :

$$\tan(\theta) = \frac{P_y}{P_z} = \frac{f\,\frac{p_v}{p_z}}{1 + v_0\,\frac{p_v}{p_z}} \tag{17}$$

using (3) and (8). In the sequel we consider $v_0 = 0$.

If $t_x \neq 0$ or $t_y \neq 0$, in this situation, we have discussed the fact that we can estimate p_z from $\bar{u}_0$ and/or $\bar{v}_0$, whereas if $t_x = t_y = 0$ we obtain a biased estimation of p_z from (15). As a consequence, we obtain a biased estimation of the slant angle as:

$$\tilde{\tan(\theta)} = \tan(\theta) \frac{1}{1 - \frac{\dot{f}}{f} \frac{1}{t_z}} \tag{18}$$

We thus may assume that if the visual system take intrinsic parameters into account as in our model it will be able to correctly perceived the plane slant in the presence of a translation, say t_y, whereas it will not if only t_z is present, unless other cues are in used. If it does not consider intrinsic parameters, then it will implicitly always assume $\dot{f} = 0$ and thus provide a biased estimation.

This may be summarised in Fig. 5 and we would like to experimentally verify this assumption.

	It does not consider intrinsic parameters	It does consider intrinsic parameters	In fact other cues are also used
$t_z \neq 0$	Slant perceived	Slant perceived	Slant perceived
$t_z \neq 0$ and $\dot{f} \neq 0$	Slant perception biased	Slant perception biased	Slant perceived
$t_z \neq 0$, $t_x \neq 0$ and $\dot{f} \neq 0$	Slant perception biased	Slant perception not biased	Slant perceived

Figure 5: The three different conditions which allows to test our hypothesis.

From the model, we have seen that in order to recover from the perturbation induced by a focal-length variation, regarding the perception of the scene structure

A stimulus to generate only motion cues.

In order to be sure to generate a stimulus in which only motion cues are present without any "semantic of the scene" we experiment motion integration using simple stimulus with dot patterns (see [25] for a discussion).

From a computational point of view, using punctual primitives such as "dots' is also mandatory in order the retinal motion field to be observable. As reviewed in [17] for instance, the motion field is in general not entirely perceived (aperture problem) with "true" images containing edges except at corners. As a consequence, most of the computer motion perception module have the structure shown in Fig. 6, detecting and tracking features such as corners in an image. Here we also can get rid of this problem because we first estimate a parameterised retinal field (here either affine or quadratic) using robust and numerically stable estimation methods which could easily combined partial observation of the retinal field. Furthermore, motion segmentation (e.g. [5]) of different rigid objects could be integrated at this stage. This intermediate step is thus very useful before computing structure and motion parameters. The reader may refer to [39] for an implementation of a computer motion perception module in the un-calibrated case.

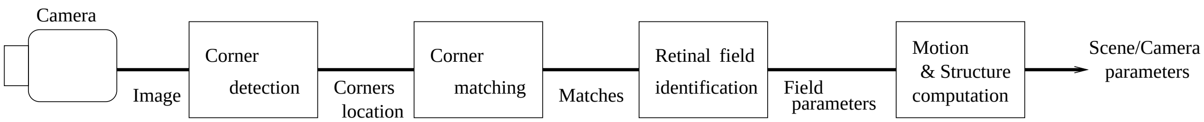

Figure 6: Implementing our model as a computer motion perception module.

Using either small or large field stimuli is an important issue, as analysed for instance in [8]. Here we have seen that in both cases the recovery of the structure and motion parameters are mainly performed using the affine part of the parameterised retinal field, so that we do not expect major differences by using a small or large stimulation field.

On the contrary a much more important issue is the fact we may propose to the user either a *passive* stimulus or an *active* stimulation, which will be investigated in a future study.

Performing a first experiment.

Description of the paradigm. A circular window is observed by the subject in a monocular and achromatic situation for a diurnal lighting. The object is a plane represented by random dots of constant density in the image, each dot having a random duration of life. The stimulus is generated at 60-80 frames/seconds on a 1600x1200 pixels screen of 40x30cm situated a 1 meter in front of the subject, thus offering a field of view of about 20 degree. Magnitude of the motions corresponds to pixels displacements of a fraction of pixel/frame.

Preliminary results, obtained on 5 voluntary subjects are given, in Fig. 7, to Fig 11. Although carefully realized by a young specialist of such psycho-metric experiments, these results are to be taken with care, more as an invitation to realized a full set of experiments in this direction, than as a final result.

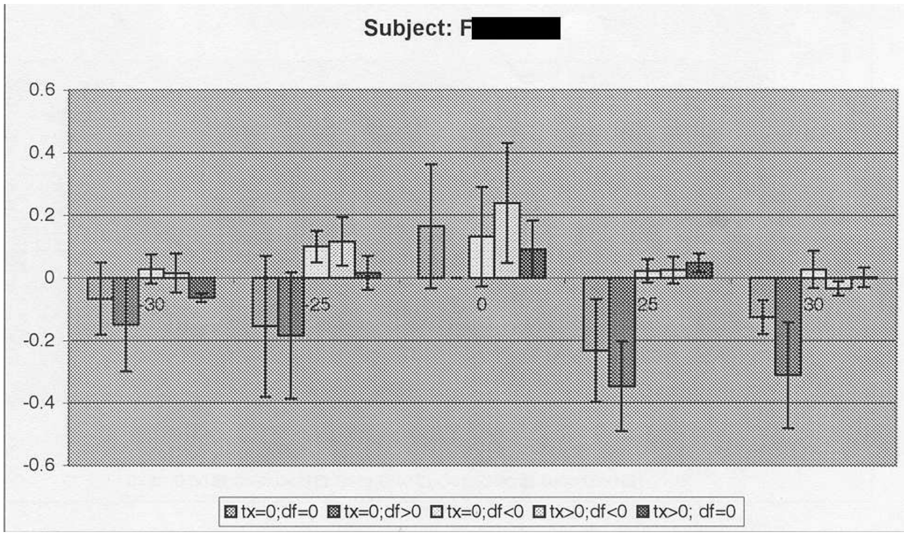

Figure 7: Experimental for the 1st subject, see text.

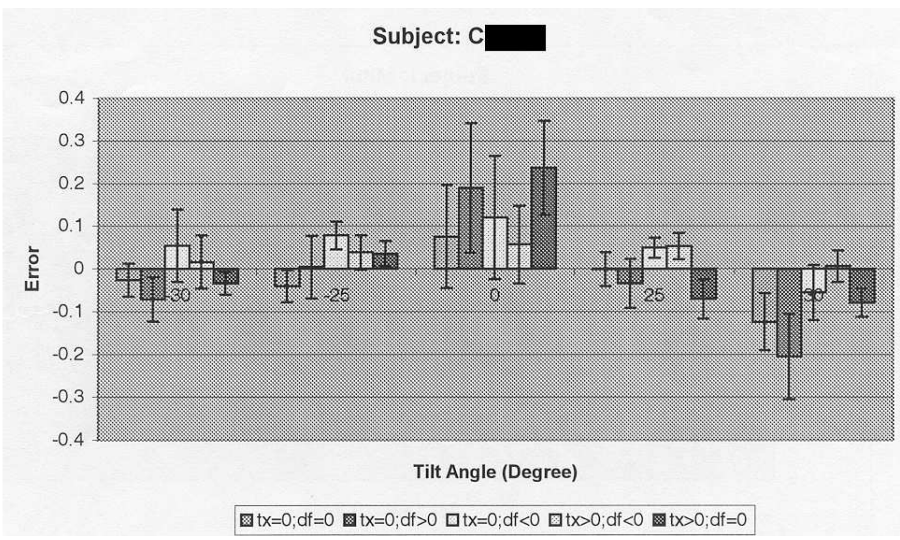

Figure 8: Experimental for the 2nd subject, see text.

In this set of experiments, 5 conditions have been tested combining forward translation and zoom variations, i.e. the 3 conditions described in Fig. 5 plus 2 control conditions:

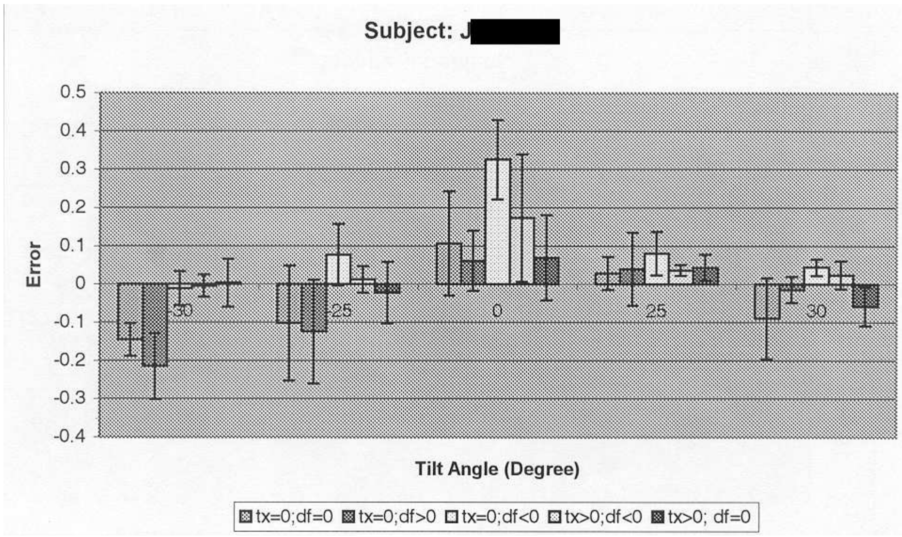

Figure 9: Experimental for the 3rd subject, see text.

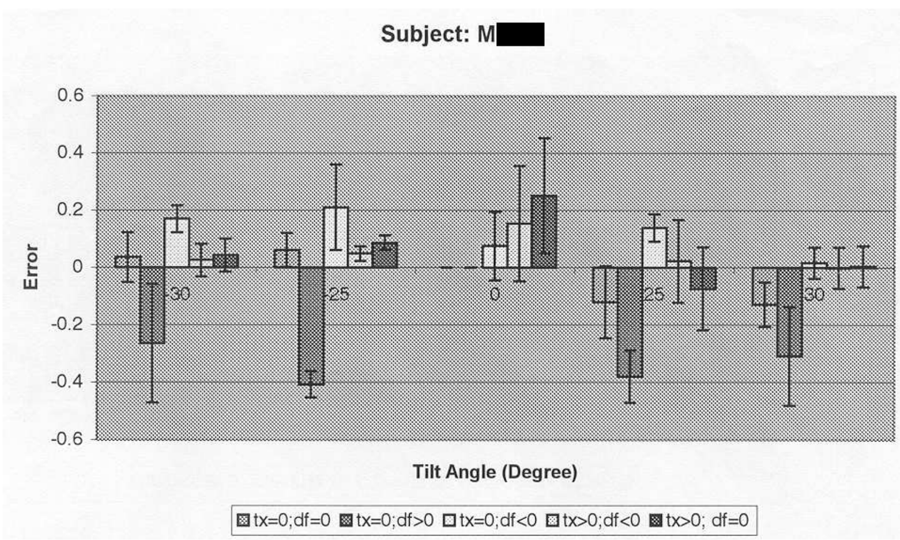

Figure 10: Experimental for the 4th subject, see text.

C1	tx $=$ 0 df $=$ 0	Simple translation in Z
C2	tx $=$ 0 df $>$ 0	A zoom in the translation direction is added
C3	tx $=$ 0 df $<$ 0	A zoom against the translation direction is added
C4	tx $>$ 0 df $>$ 0	A zoom is added but lateral translation provides another cue
C5	tx $>$ 0 df $<$ 0	A zoom is added but lateral translation provides another cue

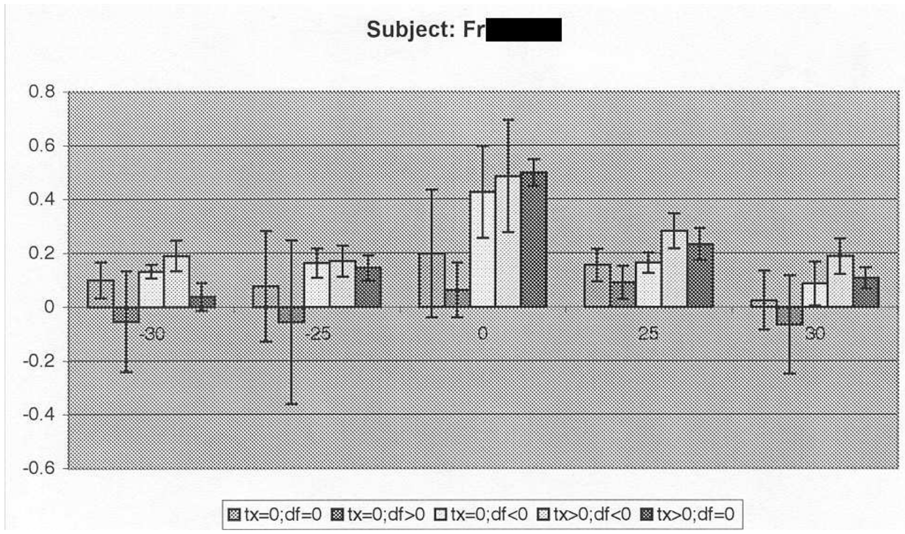

Figure 11: Experimental for the 5th subject (not taken into account), see text.

with different signs as made explicit in the figures. Each condition has been presented randomly presented 10 times and at different plane slant angles : $\Theta = \{-30, -25, 0, 25, 25\}$ degrees.

Each subject has been asked to fix the screen centre and to manipulate a small line-segment to be oriented in the direction of the perceived plane normal. The adjustment error is measured and reported.

Analysis of the results. Clearly, these results are rather random and we can not infer much from them.

We have however been able to establish a few statistical results, after exclusion of the 5th subject, which has reported not being able to perform the task and which results are not in coherence with the others. We however produce the corresponding data for completeness.

First of all, errors for fronto-parallel planes are much more erratic than others. A Student test on assuming that $(\mathcal{H}_1)$ the error for a fronto-parallel plane is not equal to the other errors is significant with a probability of $P = 0.99$. This means that, in this subjects, attempt to "find a slant", i.e. to over-estimate the plane slant.

On the other hand, we have noticed that result are asymmetric for two subjects, a focal length increase (df > 0) (i.e. in the sense of a Z translation in the front direction) induce a higher bias (when lateral translation is not present to preserve it) than a focal length decrease. A Student test on assuming that $(\mathcal{H}_1)$ the absolute value is different is not equal for both conditions is significant with a probability of $P = 0.95$.

Regarding our assumption, it is clear that a stimulus with a zoom (C2 and C3) but no lateral translation to disambiguate the estimation of the translation in Z with respect to focal length variation (as it is the case in C4 and C5). Using a Student test again we have the evidence on three subjects with a probability of $P = 0.95$ and one subject with a probability of $P = 0.99$ that these two results differ, while the last subject is not significant in any case as if it was not able to perform the experiment.

In coherence with this result, considering the 4 "correct"subjects, it is clear that a stimulus with a zoom (C2 and C3) but no lateral translation to disambiguate the estimation of the translation induces a significant bias (Student test with a probability of $P = 0.99$) than the control condition C1, whereas this could not be done when comparing conditions C3 and C4 with respect to C1 (a Student test with a weak probability of such as $P = 0.9$ fails to show a significative difference).

As a conclusion, we have been able to realize an experiment which is coherent with "It does consider intrinsic parameters" assumption developed in Fig. 5.

We are indeed very cautious with respect to these encouraging but not definitive results and the partners of this study competent in psycho-physics will carry on experiments in this direction.

Discussion

We must consider intrinsic parameters when discussing the perception of structure from motion.

Both the theoretical developments and the small experiment presented here, tend to show that usual models of structure from motion perception are wrong by starting saying "let us consider a [calibrated] visual system". On the contrary, the present developments have demonstrated that :

- considering intrinsic parameters (in the form usually considered in computer vision) and their variations does not fundamentally change the nature of the structure-from-motion equations, but allow to derive equations which are *independent* of their occurrence,
- the perception of intrinsic parameters with their variations is not independent of other motion parameters, while only combinations of them are measurable, depending on the experimental paradigm; however in any cases, the perception of scene structure parameters can be performed, *without* evaluating them explicitely,
- the human visual system seems to be able to take intrinsic parameters variations into account, during the perception of such parameters.

A step further, several discussions are opened by considering these new parameters, for instance :

- whether the fact that the eyes are performing, with a huge accuracy, pure rotations is not related not only to the need of not having parallax when putting two retinal in correspondences (for instance before and after a saccade) but also with self-calibration of the eye intrinsic parameters : the present developments show that it is plausible,
- whether the string relationships between small/large visual fields, affine/quadratic parameterisation and the perception of curvature which have been made explicit here, really correspond to what is perceived by the visual system, if so, showing that curvature is not easily perceived by a foveal motion perception whereas larger retinal fields are needed,

raising other possibilities of investigations about the structure and motion perception.

Let us now discus some of these perspectives, i.e. how we could investigate more deeply visual cognition, using such kind assumptions.

Other perceptual experiments on intrinsic parameters.

Considering the proposed paradigm we could have

- (i) distinguished a condition with a small versus a large visual field in order to verify if second order terms are, as predicted, not of great influence in this case, although this is already assessed in another context by [21, 7],
- (ii) compared with motions generated by active displacements of the subject as discussed previously, this will be treated in a near future, by our two teams,
- (iii) observed also non-planar surface to confirm that curvatures are detected by second order-terms and have no direct influence on the aspects of motion perception analysed here, as [9, 10] have implicitly verified.

A step further, many other experiments could help understanding more deeply how visual cognition manage the problems related to calibration. Let us give two examples :

The perception of the horizon Perceiving the horizon means having an affine calibration of the scene (e.g. [15, 45, 2]). With such a calibration, it is possible to perceive for instance the middle of two points or if lines are parallel. Stimulus with or without a background "at infinity" could be investigated in order to see if this helps measuring such affine cues.

Localisation with respect to a plane As discussed in this paper, without any calibration it is in theory still possible to detect the position of a plane with respect to a plane. This could be tested either by experimenting about the localisation of a plane with respect to a plane or if two points are on the size side (versus at opposite sides) with respect to a given plane.

Perturbing Euclidean visual perception. In order to be sure that the visual system would not be in a situation where it could use Euclidean or even affine cues, but only projective cues, we could either :

- generates random variations of the intrinsic parameters, which would be interpreted as spurious displacements (rotations for principal point displacement, translation in depth for focal length variation) by a visual system based on a model assuming calibration,
- increase the noise in the stimulus since we may expect the visual system to have more difficulties to use Euclidean cues, based on parameterisation with more degrees of freedom, if the input data is degraded.

These two ways to perturb Euclidean visual perception may also be used to investigate other aspects of visual perception such as structure perception (such as experiment realized in [29]) or even object recognition or discrimination.

Investigating motion/structure ambiguities. Another track is related to the perception of "oriented" projective geometry [35, 22] which could be used to explain cognitive perception of the correct orientation, for instance the fact human correctly perceived the orientation of a plane, despite an inherent ambiguity (e.g. [42]).

In relation with this orientation ambiguity is the orientation-translation ambiguity in the case of a planar stimulus, as discussed in this paper. Despite the fact that normally, the solution corresponding to the plane with tends to be the more fronto-parallel, one may try to generate other perceptions by considering stimulus which two solutions correspond to planes with the same slant, starting with a condition where translation and plane normal are parallel (no ambiguity) and increasing this ambiguity in order to see if a subject can "track" the solution corresponding to a plane slanted, etc ...

Toward higher-oder motion perception. Another important issue is the fact we propose a model based on first-order equations, only involving velocities, but not accelerations. This corresponds to a "two views" problem in the discrete case, although it is known that "three views" are needed for several aspects of un-calibrated motion perception (e.g. [40]). This aspect is to investigated :
(a) presenting to the subject, not a stimulus given as an image sequence, but as a 2, 3 or more "time slots",

(b) verifying if motion cues are not perceived differently depending on the number of views.

A step further, one should also consider the fact that the intrinsic parameters (u_0, v_0, f) in fact strongly vary spatially on the eye retina. This last consideration may indeed mark the limit of the proposed model to understand the relation between the eye intrinsic parameters and the related visual perception.

A Symbolic derivations

```
# Second order retinal motion equations
#  Here :
#   d_u0 = f diff(u_0 / f, t),
#   d_v0 = f diff(v_0 / f, t),
#   d_f = diff(f, t) / f,
#   f2 = f^2

eqs := {
 du_0 = d_u0 + p_z * t_x + r_y   - u_0 * (v_0 * r_x - u_0 * r_y) / f2,
 du_1 = p_u * t_x + d_f - p_z * t_z
         + (v_0 * r_x - 2 * u_0 * r_y + u_0 * v_0 * r_z) / f2,
 du_2 = p_v * t_x - r_z          + u_0 * (r_x - u_0 * r_z) / f2,
 du_11 = -p_u * t_z + p_uu * t_x + (r_y - v_0 * r_z) / f2,
 du_12 = -p_v * t_z + p_uv * t_x - (r_x - u_0 * r_z) / f2,
 du_22 = 0          + p_vv * t_x,

 dv_0 = d_v0 + p_z * t_y - r_x   - v_0 * (v_0 * r_x - u_0 * r_y) / f2,
 dv_1 = p_u * t_y + r_z          - v_0 * (r_y - v_0 * r_z) / f2,
 dv_2 = p_v * t_y + d_f - p_z * t_z
         + (2 * v_0 * r_x - u_0 * r_y - u_0 * v_0 * r_z) / f2,
 dv_11 = 0          + p_uu * t_y,
 dv_12 = -p_u * t_z + p_uv * t_y + (r_y - v_0 * r_z) / f2,
 dv_22 = -p_v * t_z + p_vv * t_y - (r_x - u_0 * r_z) / f2
}:

# (1) Solving in the case of a foveal motion,
```

```
eqs1 := subs(f2 = 1 / zero, zero = 0, eqs):
s1 := u -> if type(u, `=`) then op(1,u) = s1(op(2,u)) else
 collect(numer(u), {t_x, t_y, t_z}, distributed, factor) /
 collect(denom(u), {t_x, t_y, t_z}, distributed, factor) fi:

#  (1.1) eliminating structural and rotational parameters

sl11 := map(s1, leastsqrs(eqs1, {p_uu, p_uv, p_vv, r_x, r_y, r_z, d_f})):
eqs11 := op(2,eliminate(eqs1, {p_uu, p_uv, p_vv, r_x, r_y, r_z, d_f})):

sl11 := sl11 union
 map(s1, solve(map(u -> if has(u, {du_1,dv_1,du_2,dv_2}) then u fi, eqs11),
 {p_u, p_v}));
                t_x du_22 + t_y dv_22 + t_y p_v t_z
sl11 := {p_vv = -----------------------------------,
                              2      2
                           t_y  + t_x

    r_y = du_0 - d_u0 - p_z t_x, r_x = -dv_0 + d_v0 + p_z t_y,

    d_f = 1/2 dv_2 + 1/2 du_1 - 1/2 p_u t_x + p_z t_z - 1/2 p_v t_y,

    r_z = - 1/2 p_u t_y - 1/2 du_2 + 1/2 p_v t_x + 1/2 dv_1,

           t_y dv_11 + t_x du_11 + t_x p_u t_z
    p_uu = -----------------------------------,
                         2      2
                      t_y  + t_x

           t_x du_12 + t_x p_v t_z + t_y dv_12 + t_y p_u t_z
    p_uv = -------------------------------------------------,
                                2      2
                             t_y  + t_x

          (dv_1 + du_2) t_x + (dv_2 - du_1) t_y
    p_v = -------------------------------------,
```

```
                         2      2
                      t_y  + t_x

           (-dv_2 + du_1) t_x + (dv_1 + du_2) t_y
    p_u = -------------------------------------}
                         2      2
                      t_y  + t_x

eqs12 := op(2, eliminate(eqs11, {p_u, p_v})):

#  (1.2) solving with respect to the translation on the remaining equations

#  since there exists linear constraints on the translation from the equations

sl12 := {
 (dv_11 - dv_22) * t_x + (du_22 - du_11) * t_y - (du_2 + dv_1) * t_z,
  du_12 * t_y - dv_12 * t_x + (dv_2 - du_1) * t_z}:

evalb(normal(subs(eqs1, sl12)) = {0});
                                      true


map(s1, simplify(eqs12, sl12));
        2                2                                  2
{0, -t_x  du_22 t_y + t_y  t_x dv_22 + (dv_1 + du_2) t_z t_x

                                         3            3
     + (dv_2 - du_1) t_x t_z t_y + t_x  dv_22 - t_y  du_22}


# (1.3) Foveal motion, when t_x = t_y = 0

eqs13 := subs(f2 = 1 / zero, zero = 0, {t_x = 0, t_y = 0}, eqs):

map(s1, leastsqrs(eqs13, {r_x, r_y, r_z, d_f, p_u, p_v}));
           -du_12 - dv_22            -du_11 - dv_12
{p_v = 1/2 --------------, p_u = 1/2 --------------, r_y = du_0 - d_u0,
```

```
              t_z                          t_z

    r_x = -dv_0 + d_v0, r_z = - 1/2 du_2 + 1/2 dv_1,

    d_f = 1/2 du_1 + p_z t_z + 1/2 dv_2}


# (1.4) Considering a retinal rotation

eqs14 := subs({r_x = f * R_x + u_0 * R_z, r_y = f * R_y + v_0 * R_z},
  {R_x = 0, R_y = 0, R_z = r_z, f2 = f^2}, eqs):

evalb(subs({r_x = u_0 * r_z, r_y = v_0 * r_z}, eqs1) = eqs14);
                                      true


# (1.5) Considering a case without rotation

eqs15 := subs({r_x = 0, r_y = 0, r_z = 0}, eqs):
eqs15 := op(2, eliminate(eqs15, {p_uu, p_uv, p_vv, d_f})):
eqs15 := op(2, eliminate(eqs15, {p_u, p_v})):

map(s1, leastsqrs(eqs15, {p_z}));
                        (du_0 - d_u0) t_x + (dv_0 - d_v0) t_y
                 {p_z = -------------------------------------}
                                        2      2
                                     t_y  + t_x


# (2) Pure image transformation : rotation and intrinsic parameters variation

eqs2 := subs(
  {t_x = 0, t_y = 0, t_z = 0},
  {r_x = f * R_x + u_0 * R_z, r_y = f * R_y + v_0 * R_z, r_z = R_z},
  {f2 = f^2}, eqs):

# We obtain, by linear combinations of these twelve equations
```

```
eqs21 := {
#  three linear equations in u_0, v_0 and d_f
(du_2 + dv_1)   + u_0 * (du_12 + dv_22)/2 + v_0 * (du_11 + dv_12)/2,
(dv_2 - du_1)   + v_0 * (du_12 + dv_22)/2 - u_0 * (du_11 + dv_12)/2,
(dv_2 + du_1)/3 + u_0 * (du_11 + dv_12)/2 + v_0 * (du_12 + dv_22)/2 - 2/3 * d_f,
#  two linear equations in f^2, d_u0, d_v0 considering u_0, v_0 and d_f as known
du_0 - u_0 * (dv_2 - 2 * du_1 + d_f)/3 + v_0 * du_2
 - f^2 * (du_11 + dv_12)/2 - d_u0,
dv_0 + v_0 * (2 * dv_2 - du_1 - d_f)/3 + u_0 * dv_1
 - f^2 * (du_12 + dv_22)/2 - d_v0,
#  three direct equations in R_x, R_y, R_z, intrinsic parameters being known
(du_2 - dv_1)     + u_0 * (du_12 + dv_22)/2 - v_0 * (du_11 + dv_12)/2 + 2 * R_z,
(du_12 + dv_22)/2 + R_x / f,
(du_11 + dv_12)/2 - R_y / f,
#  plus four linear constraints between coefficents
dv_11,
du_22,
du_11 - dv_12,
du_12 - dv_22
}:
# as verified here :
 evalb(normal(subs(eqs2, eqs21)) = {0});
                                       true


# If the 6 affine coeffs are available only, eliminating R_x, R_y, R_z yields
eqs22 := {
u_0 * (du_0 - d_u0) - (f^2 + u_0^2) * (dv_2-2*du_1 + d_f)/3 + u_0 * v_0 * du_2,
v_0 * (dv_0 - d_v0) + (f^2 + v_0^2) * (2*dv_2-du_1 - d_f)/3 + u_0 * v_0 * dv_1,
v_0 * (du_0 - d_u0) + u_0 * (dv_0 - d_v0) + f^2 * (dv_1 + du_2) +
 v_0^2 * du_2 + u_0^2 * dv_1 + u_0 * v_0 * (dv_2 + du_1 - 2 * d_f)/3
}:
# verifying these 3 independent linear combination of the original equations :
 evalb(normal(subs(eqs2, eqs22)) = {0});
                                       true
```

```
# solving in the case of constant intrinsic parameters :
#  (a unique couple non trivial solutions is obtained)
nops(map(u -> if not has(u, f^2 = 0) then convert(u, horner) fi,
 [solve(subs({d_u0 = 0, d_v0 = 0, d_f = 0}, eqs22), {u_0, v_0, f^2})]));
                                       1


# (3) Considering the planar case

eqs3 := subs({p_uu = 0, p_uv = 0, p_vv = 0}, eqs):

# These equations can be written as
eqs31 := {
#  two equations solvable in p_u, p_v unless t_x = t_y = 0
(dv_1 + du_2) - (p_u * t_y + p_v * t_x + (u_0 * R_x - v_0 * R_y)/f),
(du_1 - dv_2) - (p_u * t_x - p_v * t_y - (v_0 * R_x + u_0 * R_y)/f),
#  with three equations implying the intrinsic parameters variation
du_0 - ((d_u0 + p_z * t_x + r_y) - u_0 * (v_0 * R_x - u_0 * R_y)/f),
dv_0 - ((d_v0 + p_z * t_y - r_x) - v_0 * (v_0 * R_x - u_0 * R_y)/f),
(du_1 + dv_2) - (2 * (d_f - p_z * T_z)
 + p_u * t_x + p_v * t_y + 3 * (v_0 * R_x - u_0 * R_y)/f),
#  and three equations to eliminate R_x, R_y, R_z
(du_12 + dv_22) / 2 + p_v * T_z + R_x / f,
(du_11 + dv_12) / 2 + p_u * T_z - R_y / f,
(dv_1 - du_2) - (p_u * t_y - p_v * t_x - (u_0 * R_x + v_0 * R_y)/f + 2 * R_z),
#  plus four linear constraints between coefficents
dv_11,
du_22,
du_11 - dv_12,
du_12 - dv_22
}:
# so that the system remains unsolved in (u_0, v_0, f) and (t_x, t_y, T_z)
# as verified here :
 evalb(normal(subs(eqs3, {
  r_x = f * R_x + u_0 * R_z,
  r_y = f * R_y + v_0 * R_z,
```

```
  r_z = R_z, t_z = T_z, f2 = f^2}, eqs31)) = {0});
                                    true


# In the case of a ``retinal rotation'' i.e. R_x = R_y = 0 we obtain
eqs32 := [
#  two equations solvable in p_u, p_v unless t_x = t_y = 0
(dv_1 + du_2) - (p_u * t_y + p_v * t_x),
(du_1 - dv_2) - (p_u * t_x - p_v * t_y),
#  then two equations solvable in (d_f - p_z * T_z) and R_z
(du_1 + dv_2) - (2 * (d_f - p_z * T_z) + p_u * t_x + p_v * t_y),
(dv_1 - du_2) - (2 * R_z               + p_u * t_y - p_v * t_x),
#  with two equations in function the principal point variation
du_0 - (d_u0 + p_z * t_x + u_0 * R_z),
dv_0 - (d_v0 + p_z * t_y - v_0 * R_z),
#  and two equations relating the plane orientation and T_z
(du_12 + dv_22) / 2 + p_v * T_z,
(du_11 + dv_12) / 2 + p_u * T_z,
#  plus four linear constraints between coefficents
dv_11,
du_22,
du_11 - dv_12,
du_12 - dv_22
]:
# as verified here :
 evalb(normal(subs(eqs3, {
  r_x = f * R_x + u_0 * R_z,
  r_y = f * R_y + v_0 * R_z,
  r_z = R_z, t_z = T_z, f2 = f^2}, {R_x = 0, R_y = 0}, {op(eqs31)})) = {0});
                                    true

# which could be solved for p_u, p_v
sl32 := map(s1, solve({op(1..2,eqs32)},{p_u, p_v}));
               (dv_1 + du_2) t_x + (dv_2 - du_1) t_y
sl32 := {p_v = -------------------------------------,
                               2      2
                            t_y  + t_x
```

```
        (-dv_2 + du_1) t_x + (dv_1 + du_2) t_y
p_u = --------------------------------------}
                          2       2
                       t_y  + t_x

# and, if either d_f - p_z * T_z = 0 or R_z = 0, for phi,
eqs33 := map(
 u -> collect(combine(simplify(u), trig), {cos(2*phi), sin(2*phi)}, factor),
 subs(sl32, {t_x = A * cos(phi), t_y = A * sin(phi)}, [op(3..4,eqs32)]));
eqs33 := [(dv_2 - du_1) cos(2 phi) + (-dv_1 - du_2) sin(2 phi) + du_1

     + 2 p_z T_z + dv_2 - 2 d_f,

    (dv_1 + du_2) cos(2 phi) + (dv_2 - du_1) sin(2 phi) - 2 R_z + dv_1 - du_2]

# as a simple trigonometric equation

quit
```

References

[1] K. Åström and A. Heyden. Continuous time matching constraints for image streams. *I.J.C.V.*, 28(1):85–96, 1998.

[2] P. Beardsley, A. Zisserman, and D. Murray. Navigation using affine structure from motion. In J.-O. Eklundh, editor, *Proceedings of the 3rd European Conference on Computer Vision*, volume 2 of *Lecture Notes in Computer Science*, pages 85–96, Stockholm, Sweden, May 1994. Springer-Verlag.

[3] D. Bondyfalat and S. Bougnoux. Imposing euclidean constraints during self-calibration processes. In R. Koch and L. V. Gool, editors, *Proceedings of SMILE Workshop on Structure from Multiple Images*, Lecture Notes in Computer Science. Springer Verlag, Lecture Notes in Computer Science, 1998.

[4] S. Bougnoux. From projective to euclidean space under any practical situation, a criticism of self-calibration. In *IEEE International Conference on Computer Vision*, pages 790–796, 1998.

[5] P. Bouthemy and E. François. Motion segmentation and qualitative dynamic scene analysis from an image sequence. *The International Journal of Computer Vision*, 10(2):157–182, Apr. 1993.

[6] P. Brand, R. Mohr, and P. Bobet. Distorsions optiques : correction dans un modèle projectif. Technical Report 1933, LIFIA–INRIA Rhône-Alpes, 1993.

[7] V. Cornilleau-Pérès and J. Droulez. The visual perception of 3d shape from self-motion and object-motion. *Vision Research*, 34:2331–2336, 1994.

[8] T. Dijkstr, V. Cornilleau-Pérès, C. Gielen, and J. Droulez. Perception of 3d shape from ego- and object-motion: comparison between small- and large-field stimuli. *Vision Research*, 35:453–462, 1995.

[9] J. Droulez. Input, assumption & optic flow processing (a model synthesis), 1998. Unpublished.

[10] J. Droulez. Optic flow analysis : A model comparison, 1998. Unpublished.

[11] J. Droulez and V. Cornilleau. Adaptive changes in perceptual responses and visuomanual coordination during exposure to visual metrical distortion. *Vision Research*, 26(11):1783–1792, 1986.

[12] R. Enciso. *Auto-Calibration des Capteurs Visuels Actifs. Reconstruction 3D Active.* PhD thesis, Université Paris XI Orsay, Dec. 1995.

[13] R. Enciso and T. Vieville. Experimental self-calibration from four views. In C. Braccini-etal, editor, *8th International Conference Image Analysis and Processing (ICIAP'95)*, volume 974 of *Lecture Notes in Computer Science*, pages 307–312, San remo, Italy, Sept. 1995. Springer.

[14] B. Espiau. Effect of camera calibration errors on visual servoing in robotics. In *3rd International Symposium on Experimental Robotics, Kyoto,japan*, 1993.

[15] O. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? In *Proceedings of the 2nd ECCV*, pages 563–578, may 1992.

[16] O. Faugeras. Non-metric representations in 3-D artificial vision. *Nature*, 1993. Submitted.

[17] O. Faugeras. *Three-Dimensional Computer Vision: a Geometric Viewpoint.* MIT Press, 1993.

[18] R. Hartley. Euclidean reconstruction from uncalibrated views. In J. Mundy and A. Zisserman, editors, *Applications of Invariance in Computer Vision*, volume 825 of *Lecture Notes in Computer Science*, pages 237–256, Berlin, Germany, 1993. Springer-Verlag.

[19] R. Hartley. Self-calibration from multiple views with a rotating camera. In J.-O. Eklundh, editor, *Proceedings of the 3rd European Conference on Computer Vision*, volume 800-801 of *Lecture Notes in Computer Science*, pages 471–478, Stockholm, Sweden, May 1994. Springer-Verlag.

[20] A. Heyden and K. Åström. Euclidean reconstruction from image sequences with varying and unknown focal length and principal point. In *Comp. Vision and Pattern Rec.*, pages 438–443. IEEE Computer Society Press, 1997.

[21] J.Droulez and V. Cornilleau Perez. Visual perception of surface curvature. the spin varaition and its physiological implications. *Biological Cybernetics*, 1991.

[22] S. Laveau and O. Faugeras. Oriented projective geometry for computer vision. In B. Buxton, editor, *Proceedings of the 4th European Conference on Computer Vision*, pages 147–156, Cambridge, UK, Apr. 1996.

[23] D. Lingrand and T. Viéville. Dynamic foveal 3D sensing using affine models. In *Proceedings of the International Conference on Pattern Recognition*, volume 1, pages 810–814, Vienna, Austria, Aug. 1996. Computer Society Press.

[24] H. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, 1981.

[25] J. Lorenceau. Motion integration with dot patterns: Effects of motion noise and structural information. *Vision Research*, 36:3415–3427, 1996.

[26] Q. Luong and T. Viéville. Canonical representations for the geometries of multiple projective views. *Computer Vision and Image Understanding*, 64(2):193–229, 1996.

[27] Q.-T. Luong and O. Faugeras. The fundamental matrix: theory, algorithms, and stability analysis. *The International Journal of Computer Vision*, 1994.

[28] S. J. Maybank and O. D. Faugeras. A theory of self-calibration of a moving camera. *The International Journal of Computer Vision*, 8(2):123–152, Aug. 1992.

[29] T. S. Meese and M. G. Harris. Computation of surface slant from optic flow: Orthogonal components of speed gradient can be combined. *Vision Research*, 37:2369–2379, 1997.

[30] R. Mohr and E. Arbogast. It can be done without camera calibration. *Pattern Recognition Letters*, 12:39–43, Mar 19 1990.

[31] K. Nakayama. Biological image motion processing: A review. *Vision Research*, 25:625–660, 1984.

[32] Y. Nomura, M. Sagara, H. Naruse, and A. Ide. Simple calibration algorithm for high-distorsion-lens camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(11):1095–1099, Nov. 1992.

[33] L. Robert and O. Faugeras. Relative 3-D positioning and 3-D convex hull computation from a weakly calibrated stereo pair. *Image and Vision Computing*, 13(3):189–197, 1995. also INRIA Technical Report 2349.

[34] A. T. Smith and R. J. Snowden. *Visual Detection of Motion.* Academic Press, 1994.

[35] J. Stolfi. *Oriented Projective Geometry, A Framework for Geometric Computations.* Academic Press, Inc., 1250 Sixth Avenue, San Diego, CA, 1991.

[36] R. Y. Tsai. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. *IEEE Journal of Robotics and Automation*, 3(4):323–344, Aug. 1987.

[37] T. Viéville. Autocalibration of visual sensor parameters on a robotic head. *Image and Vision Computing*, 12, 1994.

[38] T. Viéville, E. Clergue, R. Enciso, and H. Mathieu. Experimenting 3-D vision on a robotic head. In *Proceedings of the International Conference on Pattern Recognition*, pages 739–743, Jerusalem, Israel, Oct. 1994. Computer Society Press.

[39] T. Viéville and O. Faugeras. The first order expansion of motion equations in the uncalibrated case. *CVGIP: Image Understanding*, 64(1):128–146, July 1996.

[40] T. Viéville, O. D. Faugeras, and Q.-T. Luong. Motion of points and lines in the uncalibrated case. *The International Journal of Computer Vision*, 17(1):7–42, Jan. 1996.

[41] T. Viéville and D. Lingrand. Using specific displacements to analyze motion without calibration. *IJCV*, 31(1):5–29, 1999.

[42] T. Viéville, C. Zeller, and L. Robert. Using collineations to compute motion and structure in an uncalibrated image sequence. *The International Journal of Computer Vision*, 20(3):213–242, 1996.

[43] A. M. Waxman and S. Ullman. Surface structure and three-dimensional motion from imageflow kinematics. *Int. J. of Robot. Res.*, 4, 1985.

[44] R. G. Willson. *Modeling and Calibration of Automated Zoom Lenses.* PhD thesis, Department of Electrical and Computer Engineering, Carnegie Mellon University, 1994. CMU-RI-TR-94-03.

[45] C. Zeller and O. Faugeras. Applications of non-metric vision to some visual guided tasks. In *Proceedings of the International Conference on Pattern Recognition*, pages 132–136, Jerusalem, Israel, Oct. 1994. Computer Society Press. A longer version in INRIA Tech Report RR2308.

We are very thankful to **Olivier Faugeras** for some his powerful ideas at the origin of this work.

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles - B.P. 93 - 06902 Sophia Antipolis Cedex (France)

Unité de recherche INRIA Lorraine : Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - B.P. 101 - 54602 Villers lès Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38330 Montbonnot St Martin (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - B.P. 105 - 78153 Le Chesnay Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, B.P. 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399